Behavior learning using emotional conditioning

Guillaume Grenet, Frédéric Alexandre

Behavior learning using emotional conditioning

Guillaume Grenet and Frédéric Alexandre[1]

LORIA - INRIA, Cortex team, Nancy (France)

Abstract. We present a novel way to design a control system for a robot, using emotions as a way to produce richer internal states. We believe that using a single scalar as an evaluation of the quality of the policy and stating that the goal of the agent is to gather reward, as it is proposed by reinforcement learning, is not an appropriate granularity for creating an autonomous control system : even with a fine-tuned reward function, efficient on a specific task, it is often impractical to derive any useful knoweldge from it in order to build more flexible, neuromimetics control systems. A complete shift of paradigm is necessary for a bottom up approach : the robot is given pain and pleasure perception circuits and we examine how emotions arise and are the basis for respondant and operant conditioning. Inspired from the cerebral circuits of superior mammals responsible of this behavior, we propose an implementation in an autonomous robot using models of adaptive neural networks.

1 Motivations and emotions

As sensorimotor capabilities are given to a robot, the subsequent question is to wonder what makes the robot use these capabilities and act in the environment: What are its motivations? Reinforcement learning defines some states of the environment as goals/sub-goals (necessary steps to reach the main goal) and introduces a reward signal as a measure of the distance to those goals. This learning selects optimal policies from the anticipation and unfolding through time of the consequences of each possible action on the environment. Despite the impressive results achieved by reinforcement learning, it seems that it may be too naive to consider the reward signal as a proper way to describe goals : it does not promote rich internal states, as exploited by us, humans, when asked to describe our motivational strategies.

This latter statement is the basis of the ongoing researches that we present in this paper: inspired by a behavioral, psychological and physiological analysis of motivation in humans, a richer model of internal states in robots could endow them with more realistic motivational characteristics. Several considerations are at the basis of the choices that have been made for the internal architecture:

1. **Two different circuits related to pain and pleasure perception** are the way by which we perceive states of the world (including ourselves) as goals to be avoided or looked after. Those two signals are sufficient to describe a wide range of emotions as described in [12], for instance, *fear* can be seen as the anticipation of pain while *frustration* can be seen as the absence of an expected pleasure. While it may be argued that a positive/negative single reward signal could be enough, the use of two distinct circuits allows to deal with situations where stimuli have both positive and negative valence in a satisfactory way. It is also our purpose here to demonstrate how giving an adaptive meaning to both signals can lead to an efficient control architecture.
2. Some situations are innately "unconditioned stimuli" for which we "automatically" anticipate pain or pleasure. New neutral situations can be associated by learning and become "conditioned stimuli". This is classical (or Pavlovian) conditioning, the rules of which have been extensively modeled [13][7]. In our approach, we propose that this learning (often associated to the limbic system) **is also the basis for our discrimination capabilities** (developed in the cortical temporal areas, ideally localized between the perceptual cortical regions and the limbic system): consistently with the Rescorla-Wagner rule [11], we learn to discriminate stimuli which evoke surprise (not predicted) whereas a neutral stimulus does not deserve to be recognized. This learning gives emotional "colors" to our world. Some situations are now perceived as associated to a cost (a pain) or a benefit (a pleasure). See [8] [9] for an extensive study of fear conditioning.
3. Then, depending on present internal or external active goals (eg: *I am thirsty; I am walking through the room*), **some stimuli with emotional colors are transformed into motivations** (eg: *I have to grasp a bottle of beer and it will give me some pleasure; I have to avoid the table or it will give me pain*). Using the knoweldge of the consequences of the actions on the world, the correct actions to be performed can be selected. This latter process corresponds to operant conditioning and is clearly related to reinforcement learning. We discuss later how this could lead to less computationally expensive models.

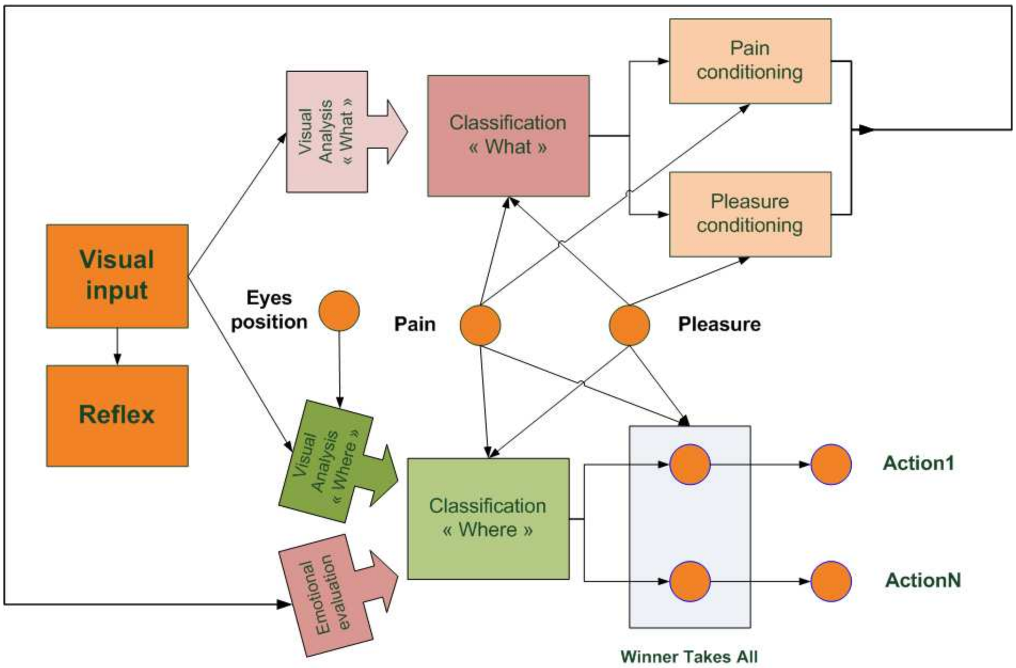

Fig. 1. Overview of the architecture

2 Architecture description

2.1 Overview

The full architecture of our proposal for a control system is shown in figure 1. First, the visual input may trigger direct reflexes after a gross analysis, which represents the innate knowledge of what is absolutely necessary to stay alive. This kind of general architecture is not without common grounds with the Per'Ac architecture introduced in [6]. Second, an adaptive visual analysis is carried along two pathways described below.

Visual processing relies on the classical *What/Where* functional analysis [14] [10] : the *What* pathway represents the analysis of what is currently shown in central vision and the *Where* pathway associates vision to actions by integrating various informations : the emotional evaluation of what has been seen in central vision, and the position of the object in the visual field (by using the angular position of the eyes). By supposing that the *What* pathway only deals with central vision, we are simplifying the analysis process and implicitly use ocular saccades to explain how the visual field is analyzed and how to contend with multiple interesting objects. This idea is similar to the embodiment level presented in [1] : in the *Where* pathway, the remaining informations from the *What* pathway will be the emotional evaluation of the object along with the angular position of the eyes that made the object be in central vision.

Considering the simple scenario where only the color of boxes scattered in the environement matters (blue could mean healthy food while red could be poisoned food), the innate knoweldge would be that when a red box is right in front of the agent, it should avoid eating it and when a box is seen somewhere in the visual field, the agent should get closer. With those two simple reflexes, it is expected to see basic survival behavior : exploration and avoidance of instant death. Over time, the agent will learn that the blue leads to pleasure (satiety) and the red will lead to pain (stomachaches), which is passive conditioning. It will also learn that when there is something that leads to pleasure on the right, it should turn right, while if it is something painful, it should go to the left to avoid it, which is operant conditioning.

As a proof of concept, we now present how each of the modules of this system has been implemented.

2.2 Visual analysis

It is not our purpose to propose here novel methods of analysis of the visual input. The *Where* pathway in our model is not purely about location, it is also a *How* pathway considering how it leads directly to action. Our

[1] Corresponding author: frederic.alexandre@loria.fr

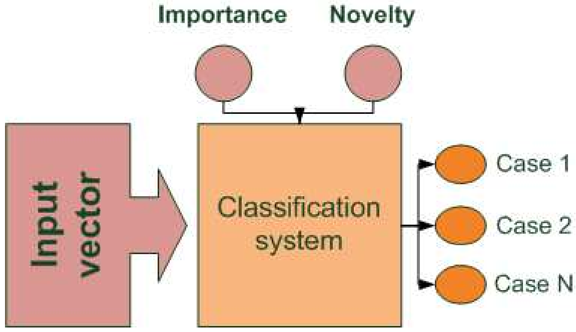

Fig. 2. Schematic representation of the classification modules

What analysis extracts shape (orientation in our experiments) and color, as in V4. Our *Where* pathway extracts the location of the object currently seen in central vision in an egocentric coordinate system, using among other informations the angular position of the eyes.

2.3 Classification modules

The *What* and *Where* classifications use the exact same implementation. The main purpose of this system is, as shown in figure 2, to classify an arbitrarily large input vector into as few cases as possible. It is an incremental neural network similar to Fritzke's Growing Neural Gas as described in [4] without the topology as it is irrelevant in this context. We decided to use incremental learning as it is clear that it is not possible to consider all possible configuration of stimuli. For instance, in the *What* pathway, the characteristics extracted from the visual input are mainly shapes and colors : the number of neurons necessary to encode each possible configuration will increase heavily as more colors or shapes are considered, even though not all configurations are relevant. This is why we have added two inputs to the classification modules : *novelty* and *importance*. The novelty of a configuration in the case of the What pathway is computed from the difference between the current emotional evaluation and the real pain and pleasure signals. In other words, if the body is feeling pleasure from eating while the emotional evaluation system is not predicting it, the novelty signal will be high. Likewise, if the emotional evaluation is predicting pain while the body is feeling pleasure, the novelty will also be high. The novelty signal is irrelevant for the Where pathway, it has been set to a high value to let the body react more quickly to objects it has already identified. In both pathways, the importance signal is the sum of the pain and pleasure signals coming from the body : if it is high, it means that something is going on and learning should take place. Those two signals are used to control the creation rate of new neurons to represent configurations of input stimuli and to control how close the prototype of the neurons will be to the input stimuli. Unlike the vigilance signal in Grossberg's ART [2], which shares many common points with our classification modules, our novelty and importance signals are tightly bound to the rest of the system and are not external parameters chosen in advance.

2.4 Conditioning modules

The pain and pleasure conditioning modules are based on Moren's amygdala model, as presented in [7] , extended to allow real-time learning by adding a Short Term Memory trace to the inputs. As explained previously, the input of the conditioning modules are not directly the visual stimuli but a small subset of configurations chosen by the classification module. Concerning the non-stationary nature of the classification, we also had to modify Moren's model to unlearn association which have lost their meaning due to the modification of the classification. Specifically, when the classes in the *what* classification module change, the association of the inputs in the conditioning modules with the emotional response decrease.

2.5 Winner Takes All

Each of the output neuron of the classification in the *where* pathway is linked with a weight between 0 and 1 to each neuron in the *Winner take all* module. The neuron of the WTA which has the highest output becomes the only one firing to control the associated action. To learn those weights, we use a mechanism very similar to Q-Learning using TD(0). Once a neuron of the ouput layer of the classification has been activated, it has a very slow decay, hence keeping a trace of the fact that it has been chosen for a long time. In the TD(0) algorithm, this neuron would encode the eligibility of the action. When pleasure or pain is felt, weights increase or decrease accordingly, hence favoring or leading to avoidance of the last chosen actions, proportional to the value of the eligibility trace.

On a side note, the reflex pathway is connected to the WTA with small constant weights so that a learned behavior can quickly but not instantaneously take over reflexes.

3 Discussion

The architecture has been simulated and is now being tested on a wheeled robot, with the scenario given in the architecture overview. We expect to observe the mentioned bootstrapping effect and autonomous learning once we have adapted all the time constants and various parameters from the simulation to the actual robot.

This architecture offers a base on which higher levels of abstraction can build upon. It is an attempt at creating a comprehensive learning and control system based on biological data and vastly supported hypothesis. In a similar way than how the *What/Where* pathways are built upon the *Reflexes* pathway, we strongly believe that it will be possible to implement higher level functionalities which will take over those reflexes when necessary, such as the executive functions of the frontal lobe [3] [5].

It is interesting to note how abstract concepts such as frustration or relief (difference between the expected pleasure/pain and the actual pleasure/pain, called importance in the classification module), fear and anticipation quickly stem from the embodiment of the learning process. Our point is not to decide whether those are actual feelings or not, but how the use of the semantics of emotions can be helpful in designing learning system in a bottom up approach.

As it is a work in progress, many improvements are expected. Currently, our main focus is on the *Where* pathway : as mentioned earlier, it is our belief that one of the ways to deal with the computational complexity, that usually comes with control systems learning by classical reinforcement learning, is to firstly only consider pertinent elements (emotionaly colored) in the environment, and secondly only consider the actions related to the current movitations. Such ways of computing are possible, and even natural, with an embodied cognition and richer internal states.

References

1. Dana H. Ballard, Mary M. Hayhoe, Polly K. Pook, and Rajesh P.N. Rao. Deictic codes for the embodiment of cognition. *Behavioural and Brain Science*, 1996.
2. G.A. Carpenter and S Grossberg. Adaptive resonance theory. In *The Handbook of Brain Theory and Neural Networks, Second Edition*, pages 87–90. MIT Press, 2003.
3. Hervé Frezza-Buet. *Un modèle de cortex pour le comportement motivé d'un agent neuromimetique autonome*. PhD thesis, Université Henry Poincaré, 1999.
4. B. Fritzke. A growing neural gas network learns topologies. *Advances in Neural Information Processing Systems*, 7:625–632, 1995.
5. Joaquin M. Fuster. *The Prefrontal Cortex : Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. Raven Press, N.Y., 1980.
6. Philippe Gaussier and Stéphane Zrehen. Complex neural architectures for emerging cognitive abilities in an autonomous system. In *Proceedings of the PerAc Conference: from Perception to Action*, pages 278–289, 1994.
7. Moren Jan. *Emotion and Learning : a computational model of the Amygdala*. PhD thesis, Lund University Cognitive Studies, 2002.
8. J.E. LeDoux. Emotion circuits in the brain. *Annu. Rev. Neurosci.*, 200:155–184, 2000.
9. J.E. LeDoux. The emotional brain, fear, and the amygdala. *Cell. Mol. Neurobiol.*, 2003.
10. A. D. Milner and M. A Goodale. Separate visual pathways for perception and action. *Trends in Neuroscience*, 1992.
11. Rescorla RA and Wagner AR. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, pages 64–99. Appleton Century Crofts, 1972.
12. Edmund T. Rolls. Precis of the brain and emotion. *Behavioral and brain sciences*, 23:177–234, 2000.
13. Nestor A. Schmajuk. *Animal Learning and Cognition : a neural network approach*. Cambridge University Press, 1997.
14. L.G. Ungerleider and M.Mishkin. Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience*, 6:414–417, 1982.